

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-206894

(43) 公開日 平成7年(1995) 8月8日

(51) Int.Cl.⁶

C 0 7 K 1/00

識別記号

庁内整理番号

F I

技術表示箇所

8318-4H

審査請求 未請求 請求項の数 6 O L (全 21 頁)

(21) 出願番号 特願平6-2461

(22) 出願日 平成6年(1994) 1月14日

(71) 出願人 393020225

アドバンスドテクノロジーインスティテュート株式会社

神奈川県横浜市緑区青葉台2丁目10番20号

(72) 発明者 窪田 毅

神奈川県横浜市緑区青葉台2丁目10番20号

アドバンスド テクノロジー インスティテュート株式会社

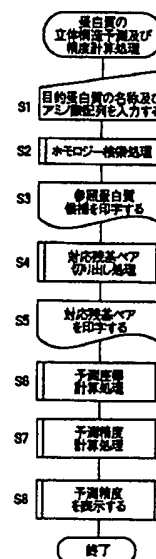
(74) 代理人 弁理士 青山 稔 (外2名)

(54) 【発明の名称】 蛋白質の立体構造の予測精度演算方法及び予測精度演算装置

(57) 【要約】

【目的】 所定の方法で予測された目的蛋白質の立体構造の予測精度をアミノ酸配列に沿って定量的に推定して演算することができる蛋白質の立体構造の予測精度演算方法及び装置を提供する。

【構成】 立体構造を予測すべき目的蛋白質との相同性の度合いが高い少なくとも1つの蛋白質を参照蛋白質に選択し、選択した参照蛋白質のアミノ酸残基の原子の座標に基づいて目的蛋白質のアミノ酸残基の原子の座標が所定の方法で予測された目的蛋白質の立体構造の予測精度を演算する方法及び装置であって、目的蛋白質に含まれる複数種類のアミノ酸の物理化学的パラメータに基づいて、参照蛋白質と目的蛋白質の併置配列において各対応するアミノ酸残基に対する相関性を示す相関関数の値を、複数種類のアミノ酸の物理化学的パラメータから予め選択された少なくとも1つの選択物理化学的パラメータについてそれぞれ計算して平均値を計算して予測精度として用いる。



【特許請求の範囲】

【請求項1】 アミノ酸残基の原子の座標が既知の蛋白質のうち、立体構造を予測すべき目的蛋白質との相同性の度合いが高い少なくとも1つの蛋白質を参照蛋白質に選択し、上記選択した参照蛋白質のアミノ酸残基の原子の座標に基づいて上記目的蛋白質のアミノ酸残基の原子の座標が所定の方法で予測された目的蛋白質の立体構造の予測精度を演算する蛋白質の立体構造の予測精度演算方法であって、

上記目的蛋白質に含まれる複数種類のアミノ酸の物理化学的パラメータに基づいて、上記参照蛋白質と上記目的蛋白質の併置配列において各対応するアミノ酸残基に対する相関性を示す相関関数の値を、上記複数種類のアミノ酸の物理化学的パラメータから予め選択された少なくとも1つの選択物理化学的パラメータについてそれぞれ計算し、上記計算した相関関数の値を平均して計算された相関関数の平均値を上記予測された目的蛋白質の立体構造の予測精度として用いることを特徴とする蛋白質の立体構造の予測精度演算方法。

【請求項2】 上記アミノ酸残基の原子は α 炭素原子であることを特徴とする請求項1記載の蛋白質の立体構造の予測精度演算方法。

【請求項3】 上記選択物理化学的パラメータは、極性度と、偏比容と、ターン形成度と、 α アミノ基のpK値と、 α カルボキシル基のpK値と、突然変異度とのうちの少なくとも1つを含むことを特徴とする請求項1又は2記載の蛋白質の立体構造の予測精度演算方法。

【請求項4】 アミノ酸残基の原子の座標が既知の蛋白質のうち、立体構造を予測すべき目的蛋白質との相同性の度合いが高い少なくとも1つの蛋白質を参照蛋白質に選択し、上記選択した参照蛋白質のアミノ酸残基の原子の座標に基づいて上記目的蛋白質のアミノ酸残基の原子の座標が所定の方法で予測された目的蛋白質の立体構造の予測精度を演算する蛋白質の立体構造の予測精度演算装置であって、

上記目的蛋白質に含まれる複数種類のアミノ酸の物理化学的パラメータに基づいて、上記参照蛋白質と上記目的蛋白質の併置配列において各対応するアミノ酸残基に対*

*する相関性を示す相関関数の値を、上記複数種類のアミノ酸の物理化学的パラメータから予め選択された少なくとも1つの選択物理化学的パラメータについてそれぞれ計算する第1の計算手段と、

上記第1の計算手段によって計算された相関関数の値を平均して上記相関関数の平均値を計算する第2の計算手段とを備え、

上記第2の計算手段によって計算された上記相関関数の平均値を、上記予測された目的蛋白質の立体構造の予測精度として用いることを特徴とする蛋白質の立体構造の予測精度演算装置。

【請求項5】 上記アミノ酸残基の原子は α 炭素原子であることを特徴とする請求項4記載の蛋白質の立体構造の予測精度演算装置。

【請求項6】 上記選択物理化学的パラメータは、極性度と、偏比容と、ターン形成度と、 α アミノ基のpK値と、 α カルボキシル基のpK値と、突然変異度とのうちの少なくとも1つを含むことを特徴とする請求項4又は5記載の蛋白質の立体構造の予測精度演算装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、蛋白質工学などのバイオテクノロジーの分野における蛋白質の立体構造の予測精度演算方法及び予測精度演算装置に関する。

【0002】

【従来の技術】 蛋白質とは、グルタミン酸やリジンなど表1に示す20種類のアミノ酸がペプチド結合して、例えば、メチオニン-グルタミン酸-グリシン-リジン-プロリン-……のように数十乃至数百個のアミノ酸が繋がって構成されたものであり、これをアミノ酸配列又は蛋白質の1次構造と呼んでおり、当該アミノ酸配列は、蛋白質の最も本質的な情報、特にその立体構造や機能に関する情報を有する。このように隣接する同士のアミノ酸がペプチド結合して蛋白質を構成する単位を特にアミノ酸残基と呼んでいる。

【0003】

【表1】

アミノ酸の名称	3文字コード	1文字コード
グリシン	GLY	G
アラニン	ALA	A
セリン	SER	S
システイン	CYS	C
メチオニン	MET	M
リシン	LYS	K
バリン	VAL	V
トレオニン	THR	T
イソロイシン	ILE	I

ロイシン	LEU	L
アスパラギン酸	ASP	D
アスパラギン	ASN	N
グルタミン酸	GLU	E
グルタミン	GLN	Q
アルギニン	ARG	R
プロリン	PRO	P
ヒスチジン	HIS	H
フェニルアラニン	PHE	F
チロシン	TYR	Y
トリプトファン	TRP	W

【0004】一般に、蛋白質のアミノ酸配列を決定することは比較的容易であり、現在までに約2万種類にのぼる蛋白質のアミノ酸配列が決定され、データベース化されている。一方、これらの蛋白質の立体構造については、従来、例えば、X線結晶解析又はNMRを用いた解析によって、蛋白質を構成するアミノ酸残基の各原子に対してその空間座標(x, y, z)が求められている。以下、この空間座標の蛋白質のデータベースをPDB (Protein Data Bank) という。従って、これらの立体構造が既知である蛋白質から20種類のアミノ酸残基間の相対距離に関する統計データを計算することができる。

【0005】一方、近年、上記X線解析によって決定された蛋白質の立体構造のデータの増加に伴って、予測目的とする蛋白質(以下、目的蛋白質という。)のアミノ酸配列と類似の、すなわち相同性(ホモロジー)がある立体構造が既知である参照とすべき蛋白質(以下、参照蛋白質という。)を見つけ出す可能性が高くなりつつある。このような状況のもとで、参照蛋白質の立体構造を参照することによって、目的蛋白質の立体構造を予測するというホモロジーモデリングの手法が提案されている。すなわち、アミノ酸配列が類似していれば、立体構造もよく似ているであろうという仮定に基づいている。

【0006】上記ホモロジーモデリングの手法の一例として、参照蛋白質を構成する主要な原子であってアミノ残基の中心を成す α 炭素原子(以下、 C_α 原子という。)の3次元座標を用いる方法が提案されている。当該 C_α 原子の絶対座標は、蛋白質分子の向きや位置に依存するために、上記3次元座標の計算がたいへん複雑になるという問題点があった。

【0007】一方、クリッペン(G.M.Crippen)とヘーベル(T.Havel)は C_α 原子間の相対距離からメトリック行列を求め、この固有値問題を解くことにより、大きい順で上位3つの固有値から C_α 原子の3次元座標を求めるという、いわゆる距離の幾何学(distance geometry)による方法を提案している。しかしながら、かかる方法では得られた蛋白質が実際のものより縮小されるといふ問題点があった。

【0008】この問題点を解決するために、参照蛋白質の2つの C_α 原子間の相対距離を用いて、目的蛋白質の立体構造を予測する方法が、例えば、輪湖博(Hiroshi Wako)とハロルド・エイ・シェラガ(Harold A.Scheraga)による次の文献において提案されている(以下、輪湖・シェラガの方法という。)

"Distance-Conatstraint Approach to Protein Folding.

20 I. Statistical Analysis of Protein Conformations in Terms of Distances Between Residues", Journal of Protein Chemistry, Vol.1, No.1, pp5-45, 1982 (以下、文献1という。)

【0009】しかしながら、文献1において開示された輪湖・シェラガの方法においては、X線結晶解析及びNMRを用いた解析によって得られた参照蛋白質の立体構造に基づいた平均的な統計データを使用するために、予測したい実際の蛋白質の座標と大きなずれを生じ、実際の立体構造を予測することができなかった。

30 【0010】この問題点を解決するために、本発明者は、特願平5-252895号において、蛋白質の立体構造の予測演算方法及び予測演算装置を提案した。当該演算方法及び演算装置は、アミノ酸残基の原子の座標が既知の蛋白質のうち、立体構造を予測する目的蛋白質との相同性の度合いが高い少なくとも1つの蛋白質を参照蛋白質に選択し、上記選択した参照蛋白質のアミノ酸残基の原子の座標に基づいて上記目的蛋白質のアミノ酸残基の原子の座標を予測演算する蛋白質の立体構造の予測演算方法であって、上記参照蛋白質と上記目的蛋白質との間のアミノ酸配列を比較することによって相同な複数の残基ペアを切り出し、上記切り出した複数の残基ペアに基づいて上記参照蛋白質における互いに残基番号が異なる各残基の原子間の相対距離を計算して上記目的蛋白質の対応する第1の相対距離に設定し、所定の目的蛋白質の初期の立体構造の座標に基づいて上記初期の立体構造における互いに残基番号が異なる各残基の原子間の第2の相対距離を計算し、上記目的蛋白質における互いに残基番号が異なる各残基の原子間に関する上記第1の相対距離とそれに対応する上記第2の相対距離との差の二乗の和が最小になるように最適化を行って上記目的蛋白質

質のアミノ酸残基の原子の座標を予測演算することを特徴としている。

【0011】

【発明が解決しようとする課題】ところで、蛋白質の立体構造を、例えば、本発明者の上記予測演算方法等の、ホモロジーモデリングなどの手法により予測した場合、X結晶解析などによる実際の立体構造からのズレを示す偏差 (deviation) は、アミノ酸配列に沿って一様ではないことが知られている。通常、立体構造を予測すべき目的蛋白質は、X結晶解析などの結果が知られていないので、アミノ酸配列の情報から局所的な予測精度を推定することは困難である。一方、目的蛋白質のアミノ酸配列と相同な (類似な) 配列を有する参照蛋白質が見つかった場合は、例えばホモロジーモデリングなどの手法により立体構造が予測されるが、一般的に、アミノ酸配列が類似していれば立体構造も良く似ていることが期待されているということは、アミノ酸配列の情報から局所的な予測精度の推定の可能性を示唆するものである。

【0012】しかしながら、従来の技術では、相同性の程度がアミノ酸配列に沿って定量的且つ立体構造を反映するように定義されていないために、アミノ酸配列に沿った立体構造の局所的な予測精度を定量的に推定することができなかった。すなわち、相同性の定義がアミノ酸配列に沿って、定量的且つある領域のアミノ酸配列の相同性が高ければその領域の立体構造もよく似ているようにはなされていないという問題点があった。

【0013】この問題点を解決するためには、この定量的且つ立体構造を反映していない相同性の定義の問題点については、本発明者である窪田綾 (Yasushi Kubota) らが相関関数を用いることで解決できることを次の文献で示している (以下、窪田らの方法という。)

“Correspondence of Homologies in Amino Acid Sequence and Tertiary Structure of Protein Molecules”, Biochimica et Biophysica Acta, Vol. 701, pp242-252 (1982) (以下、文献2という。)

【0014】しかしながら、文献2において開示された窪田らの方法においては、2つのアミノ酸配列を2次元的にマトリックスの形で展開しているために、アミノ酸配列に沿って1次元的に相同性を定量的に表現できないという難点があった。

【0015】本発明の目的はかかる問題点を解決し、所定の方法で予測された目的蛋白質の立体構造の予測精度をアミノ酸配列に沿って定量的に推定して演算することができる蛋白質の立体構造の予測精度演算方法及び予測精度演算装置を提供することにある。

【0016】

【課題を解決するための手段】本発明に係る請求項1記載の蛋白質の立体構造の予測演算方法は、アミノ酸残基の原子の座標が既知の蛋白質のうち、立体構造を予測すべき目的蛋白質との相同性の度合いが高い少なくとも1

つの蛋白質を参照蛋白質に選択し、上記選択した参照蛋白質のアミノ酸残基の原子の座標に基づいて上記目的蛋白質のアミノ酸残基の原子の座標が所定の方法で予測された目的蛋白質の立体構造の予測精度を演算する蛋白質の立体構造の予測精度演算方法であって、上記目的蛋白質に含まれる複数種類のアミノ酸の物理化学的パラメータに基づいて、上記参照蛋白質と上記目的蛋白質の併置配列において各対応するアミノ酸残基に対する相関性を示す相関関数の値を、上記複数種類のアミノ酸の物理化学的パラメータから予め選択された少なくとも1つの選択物理化学的パラメータについてそれぞれ計算し、上記計算した相関関数の値を平均して計算された相関関数の平均値を上記予測された目的蛋白質の立体構造の予測精度として用いることを特徴とする。また、請求項2記載の蛋白質の立体構造の予測精度演算方法は、請求項1記載の蛋白質の立体構造の予測精度演算方法において、上記アミノ酸残基の原子は α 炭素原子であることを特徴とする。さらに、請求項3記載の蛋白質の立体構造の予測精度演算方法は、請求項1又は2記載の蛋白質の立体構造の予測精度演算方法において、上記選択物理化学的パラメータは、極性度と、偏比容と、ターン形成度と、 α アミノ基のpK値と、 α カルボキシル基のpK値と、突然変異度とのうちの少なくとも1つを含むことを特徴とする。

【0017】本発明に係る請求項4記載の蛋白質の立体構造の予測演算装置は、アミノ酸残基の原子の座標が既知の蛋白質のうち、立体構造を予測すべき目的蛋白質との相同性の度合いが高い少なくとも1つの蛋白質を参照蛋白質に選択し、上記選択した参照蛋白質のアミノ酸残基の原子の座標に基づいて上記目的蛋白質のアミノ酸残基の原子の座標が所定の方法で予測された目的蛋白質の立体構造の予測精度を演算する蛋白質の立体構造の予測精度演算装置であって、上記目的蛋白質に含まれる複数種類のアミノ酸の物理化学的パラメータに基づいて、上記参照蛋白質と上記目的蛋白質の併置配列において各対応するアミノ酸残基に対する相関性を示す相関関数の値を、上記複数種類のアミノ酸の物理化学的パラメータから予め選択された少なくとも1つの選択物理化学的パラメータについてそれぞれ計算する第1の計算手段と、上記第1の計算手段によって計算された相関関数の値を平均して上記相関関数の平均値を計算する第2の計算手段とを備え、上記第2の計算手段によって計算された上記相関関数の平均値を、上記予測された目的蛋白質の立体構造の予測精度として用いることを特徴とする。また、請求項5記載の蛋白質の立体構造の予測精度演算装置は、請求項4記載の蛋白質の立体構造の予測精度演算装置において、上記アミノ酸残基の原子は α 炭素原子であることを特徴とする。さらに、請求項6記載の蛋白質の立体構造の予測精度演算装置は、請求項4又は5記載の蛋白質の立体構造の予測精度演算装置において、上記選

物物理化学的パラメータは、極性度と、偏比容と、ターン形成度と、 α アミノ基のpK値と、 α カルボキシル基のpK値と、突然変異度とのうちの少なくとも1つを含むことを特徴とする。

【0018】

【実施例】以下、図面を参照して本発明に係る実施例について説明する。図1は、本発明に係る一実施例である蛋白質の立体構造の予測演算及び予測精度演算装置の構成を示すブロック図である。

【0019】本実施例の蛋白質の立体構造の予測演算及び予測精度演算装置で用いる方法は、大きく分けて以下のステップを有することを特徴としている。

(a) PDBに基づいて、目的蛋白質のアミノ酸配列と類似する配列を有しかつ立体構造が既知の参照蛋白質を検索し、対応する相同な残基ペアが全体に占める割合を示す相同性の度合いが高い、例えば30%以上である参照蛋白質候補のうち相同性の度合いが最大（又は比較的高い）である蛋白質を参照蛋白質に選択する。

(b) 選択された参照蛋白質と目的蛋白質との間のアミノ酸配列を比較して、5残基以上の挿入又は欠失の無い対応する相同な複数の残基ペアを切り出す。

(c) 切り出された複数の残基ペアに基づいて、参照蛋白質に対応する複数の相同な領域間における互いに残基番号が異なる各残基の原子間の相対距離を計算して、それを目的蛋白質の対応する相対距離 $\langle d_{ij} \rangle$ として設定する。

(d) 所定の乱数のタネIRに基づいて乱数を発生させて乱数値に比例する値を目的蛋白質の初期の立体構造（以下、初期構造という。）の座標値として設定した後、これに基づいて上記目的蛋白質の初期構造における上記複数の相同な領域間における互いに残基番号が異なる各残基の原子間の相対距離に対応する相対距離 d_{ij} を計算する。

(e) 例えば準ニュートン法を用いて、上記目的蛋白質における互いに残基番号が異なる各残基の原子間に関する上記相対距離 d_{ij} と上記相対距離 $\langle d_{ij} \rangle$ との差の二乗の和が最小となるように、すなわち、後述する数1で示す目的蛋白質の関数Fの関数値が最小になるように関数Fの最適化を行い、関数Fの変数である目的蛋白質の C_{α} 原子の予測座標を計算する。

(f) アミノ酸配列を、例えば極性度、偏比容、ターン形成度、 α アミノ基のpK値、 α カルボキシル基のpK値、突然変異度などのような、表1の20種類のアミノ酸に固有の物理化学的パラメータpを使って数列化し、目的蛋白質Xと参照蛋白質Yとの間のアミノ酸配列の対応関係を示す併置配列（alignment）において、目的蛋白質Xの残基番号iと、その残基番号iに対応する参照蛋白質Yの残基番号jとの間の相関関数 $C_p(i)$ を後述の数5で定義して計算し、さらに、雑音軽減のため信号対雑音比S/Nの向上するために、後述の数6に示す

ように上記相関関数 $C_p(i)$ について複数のパラメータpに関する平均値 $\langle C(i) \rangle$ を計算することにより、予測精度を推定して演算する。

【0020】まず、目的蛋白質の立体構造の予測演算及び予測精度演算に関する基本的な原理について説明する。この演算において、1つの目的蛋白質は C_{α} 原子のみによって代表される残基によってモデル化されるものとする。従って、自然数M個の残基により構成される1つの目的蛋白質の立体構造はM個のポイントによって表される。ここで、M個のポイントの空間的な位置からなる立体構造は、次の数1に示す関数を最適化することによって得ることができる。

【0021】

$$\text{【数1】 } F = \sum_{i < j} w_{ij} \cdot (d_{ij} - \langle d_{ij} \rangle)^2 / m$$

【0022】関数Fは、上記数1に示すように $i < j$ の条件のもとで互いに残基番号が異なる各残基の原子間でのiとjに関する和で表される。ここで、1個のポイントをXYZの直交座標系を用いて座標を表わすと、関数Fは次の数2に示すように、3M個の座標変数の関数となる。

【0023】

$$\text{【数2】 } F = F(x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_M, y_M, z_M)$$

【0024】上記数1において、mは目的蛋白質と参照蛋白質との間で対応する残基ペアの全体の数M(M-1)/2である。また、 w_{ij} は重み係数であり、従来の輪湖・シェラガの方法においては平均的な統計データに基づいた重み係数を用いるが、本実施例においては常時 $w_{ij} = 1$ としている。さらに、 d_{ij} はi番目の残基とj番目の残基との間の相対距離であり、 $\langle d_{ij} \rangle$ は、従来の輪湖・シェラガの方法においてはPDBに基づいて多数の蛋白質から得られた平均的な統計データによる対応するi番目とj番目の残基ペアに割り当てられた相対距離であるが、本実施例においては、上述のように目的蛋白質と相同性の度合いが最大であるように選択された立体構造が既知の参照蛋白質についての上記目的蛋白質に対応するi番目とj番目の残基ペアに割り当てられた相対距離である。

【0025】アミノ酸配列が与えられると、数1の関数Fが一意的に決定されるので、本実施例においては、所定の乱数のタネIRを与えて乱数を発生させてその乱数値に比例する値を座標値とした目的蛋白質の初期構造におけるi番目の残基とj番目の残基との間の相対距離を関数Fの d_{ij} に代入して、例えば準ニュートン法を用いて、数1の関数Fが最小になるように関数Fを最適化し、最適化された関数Fを用いて目的蛋白質の C_{α} 原子の座標を計算する。すなわち、目的蛋白質の初期構造におけるi番目の残基とj番目の残基との間の相対距離と、参照蛋白質についての上記目的蛋白質に対応するi

番目とj番目の残基ペアに割り当てられた相対距離との差の二乗の和を最小にするような、目的蛋白質の座標変数 $x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_M, y_M, z_M$ を求めて、目的蛋白質の立体構造を求めている。なお、挿入又は欠失があったり、あるいは相同性の無い領域の残基には従来の方法の統計データの相対距離を用いる。

【0026】一方、従来の輪湖・シェラガの方法においては、PDBに基づいて多数の蛋白質から得られた平均的な統計データによる対応するi番目とj番目の残基ペアに割り当てられた相対距離 $\langle d_{ij} \rangle$ を用いた場合、非常に粗いデータであるので、関数Fから得られた目的蛋白質の立体構造は実際のそれと程遠いものとなる。

【0027】本実施例において、所定の乱数のタネIRを与えて乱数を発生させてその乱数値に比例する値を座標値とした目的蛋白質の初期構造を求めるときに、蛋白質の統計学的データに基づいて14個の球状蛋白質の線形回帰分析により経験的に得られた蛋白質の旋回半径(radius of gyration) R_g (Å)を示す次の数3と、当該初期構造の3M個の座標 $(x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_M, y_M, z_M)$ (以下、その符号を代表して q_{lm} とし、ここで、 $l=1, 2, 3; m=1, 2, \dots, M$ である。)を示す次の数4を用いる。

【0028】

【数3】 $\log R_g = 0.824 + 0.367 \cdot \log M$

【数4】 $q_{lm} = 2 \times R_g \times \text{rand}(IR)$

【0029】ここで、 $\text{rand}()$ は、乱数のタネを初期値としたときに乱数値 $(0 \leq \text{乱数値} < 1)$ を発生する公知の乱数発生関数である。上記数4を用いて、目的蛋白質の初期構造の座標を実際の座標により近似させるために、上記乱数値 $\text{rand}(IR)$ に蛋白質の旋回半径 R_g をかけた後、直径の座標とするためにそれを2倍して、当該初期構造の座標値とする。

【0030】さらに、上記の予測方法で予測された目的蛋白質の立体構造の予測精度を、アミノ酸配列に沿って推測して計算する方法について詳述する。アミノ酸配列は表1に示す1文字コードを用いる“記号列”で表されるため、従来の一般的に用いられる相同性は、アミノ酸残基の物理的及び化学的性質の類似性、例えばイソロイシン(I)とロイシン(L)とが類似し、グルタミン酸*40

極性度(Grantham, 1974)

8.10, 5.50, 13.00, 12.30, 5.20, 9.00, 10.40, 5.20, 11.30, 4.90,
5.70, 11.60, 8.00, 10.50, 10.50, 9.20, 8.60, 5.90, 5.40, 6.20

【0033】

※ ※ 【表3】

偏比容(Cohn and Edsall, 1943)

0.75, 0.61, 0.60, 0.66, 0.77, 0.64, 0.67, 0.90, 0.82, 0.90,

* (E)とアスパラギン酸(D)とが類似しているなどの類似性に基づいた、いわば定性的な方法によるものである。本発明の当該方法はアミノ酸配列にあるパラメータで数値化して後述の数5を用いて目的蛋白質と参照蛋白質との相関関数を求めるため、 $-1 \leq \text{相同性の程度} \leq 1$ のように定量的に定義する。すなわち、当該方法は、目的蛋白質と参照蛋白質との間の相同性の定義を、アミノ酸配列に沿って定量的且つ立体構造を反映するように行うことにより、相同性の高低から立体構造の局所的な予測精度の推定を行おうというものである。

【0031】アミノ酸配列は、公知の通り、例えば、次に示すような、極性度、偏比容、ターン形成度、 α アミノ基のpK値、 α カルボキシル基のpK値、突然変異度などのような、表1に示す20種類のアミノ酸に固有の物理化学的パラメータpを使って表2乃至表7に示すように、数値化することができる。なお、表2乃至表7では、表1に示す20種類のアミノ酸の1文字コードのアルファベット順(A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y)に数値化しており、表2乃至表7のそれぞれの()内の2つは、各パラメータの数値を提示した論文の著者名と、その発表年である。

(a) 極性度(polarity): アミノ酸残基が蛋白質表面に露出する度合い。

(b) 偏比容(partial specific volume): 単位質量のアミノ酸を無限量の溶液に溶かしたときの溶液の容積変化。

(c) ターン形成度(propensity to form reverse turn): アミノ酸残基が蛋白質内で折れ曲がる度合い。

(d) α アミノ基のpK値(pK-N): アミノ酸の α アミノ基の解離定数Kの対数に負の符号を付けた値、すなわち $-\log K$ 。

(e) α カルボキシル基のpK値(pK-C): アミノ酸の α カルボキシル基の解離定数Kの対数に負の符号を付けた値、すなわち $-\log K$ 。

(f) 突然変異度(relative mutability): アミノ酸残基の相対的な置換頻度。

【0032】

【表2】

0.75, 0.61, 0.76, 0.67, 0.70, 0.68, 0.70, 0.86, 0.74, 0.71

【0034】

* * 【表4】

ターン形成度 (Levitt, 1978)

0.77, 0.81, 1.41, 0.99, 0.59, 1.64, 0.68, 0.51, 0.96, 0.58,
0.41, 1.28, 1.91, 0.98, 0.88, 1.32, 1.04, 0.47, 0.76, 1.05

【0035】

※10※ 【表5】

αアミノ基のpK値 (Sober, 1970)

9.69, 8.35, 9.60, 9.67, 9.18, 9.78, 9.17, 9.68, 9.18, 9.60,
9.21, 8.80, 10.64, 9.13, 8.99, 9.21, 9.10, 9.62, 9.44, 9.11

【0036】

★ ★ 【表6】

αカルボキシル基のpK値 (Sober, 1970)

2.34, 1.92, 1.88, 2.10, 2.16, 2.35, 1.82, 2.36, 2.16, 2.36,
2.28, 2.02, 1.95, 2.17, 1.82, 2.19, 2.09, 2.32, 2.43, 2.20

【0037】

☆ ☆ 【表7】

突然変異度 (Dayhoff, 1978)

1.00, 0.20, 1.06, 1.02, 0.41, 0.49, 0.66,
0.96, 0.56, 0.40,
0.94, 1.34, 0.56, 0.93, 0.65, 1.20, 0.97,
0.74, 0.18, 0.41

【0038】そこで、目的蛋白質Xと参照蛋白質Yとの間のアミノ酸配列の対応関係を示す併置配列において、
目的蛋白質Xの残基番号iと、その残基番号iに対応する参照蛋白質Yの残基番号jとの間の相関性を示す相関◆

【0039】

【数5】

$$C_p(i) = \frac{\sum_{l=-k}^k (x_p(i+l) - \langle c_p \rangle) (y_p(j+l) - \langle c_p \rangle)}{\sqrt{[\sum_{l=-k}^k (x_p(i+l) - \langle c_p \rangle)^2] [\sum_{l=-k}^k (y_p(j+l) - \langle c_p \rangle)^2]}}$$

【0040】ここで、 $x_p(m)$ および $y_p(n)$ はそれぞれ、目的蛋白質Xの残基番号mと、参照蛋白質Yとの残基番号nにおけるアミノ酸残基に対する物理化学的パラメータpの値であり、また $\langle c_p \rangle$ は例えば表1の20種類のアミノ酸でのパラメータpの平均値であり、kとしては例えば5を与える。なお、挿入又は欠失などのため対応する残基ペアの無い箇所については相関関数の値を0とおく。

【0041】目的蛋白質Xと参照蛋白質Yをある1つの物理化学的パラメータで数列化し、k=5の場合は11残基のセグメント(断片)に対して図5のような対応関

40 係で上記数5を用いて相関関数 $C_p(i)$ の値を求める。なお、数5の数列和 Σ は $l=-k$ から $l=+k$ であるが、例えば $k=5$ のとき、以下のように相関関数 $C_p(i)$ の値を求める。

(a) $i=1$ (N末端) の場合は、図5の31に示すように、目的蛋白質Xの残基番号 $i, i+1, \dots, i+5$ に対してそれぞれ、参照蛋白質Yの残基番号 $j, j+1, \dots, j+5$ の6残基との間の相関関数 $C_p(i)$ の値が計算される。

(b) $i=2$ の場合は、図5の32のように、目的蛋白質Xの残基番号 $i-1, i, i+1, \dots, i+5$ に対し

てそれぞれ、参照蛋白質Yの残基番号 $j-1, j, j+1, \dots, j+5$ の7残基との間の相関係数 $C_p(i)$ の値が計算される。(c) i が最後の残基 (C末端) の場合はその逆に、目的蛋白質Xの残基番号 $i-5, \dots, i-1, i$ に対してそれぞれ、参照蛋白質Yの残基番号 $j-5, \dots, j-1, j$ の6残基との相関係数 $C_p(i)$ の値が計算される。

【0042】さらに、本発明では、雑音を除去しS/N比の向上させるために、上記計算された相関関数 $C_p(i)$ について、複数の物理化学的パラメータ p に関する平均値 (以下、相関関数の平均値という。) $\langle C(i) \rangle$ を次の数6で表わす。

【0043】

【数6】

$$\langle C(i) \rangle = \frac{1}{n} \sum_{p=1}^n C_p(i)$$

【0044】従って、本発明では、立体構造を反映し且つできるだけ独立な、すなわち互いに相関がない、例えば上記6つの物理化学的パラメータに対する上記数6を用いて (この場合 $n=6$)、アミノ酸配列に沿った1次元的な相同性の度合いを示す相関係数を計算する。以下、目的蛋白質がエラスターゼであって参照蛋白質がキモトリプシノーゲンであるときの例について説明する。

【0045】図6は、目的蛋白質であるエラスターゼの上記予測方法で予測された立体構造 (実線) とX線結晶解析によって計測された立体構造 (点線) を最適に重ね合せたものを、ステレオ図で示したものである。また、図7は、図6に基づいて、目的蛋白質エラスターゼの立体構造において、予測された当該立体構造の C_α 原子の位置について、対応するX線結晶解析による立体構造の C_α 原子からの偏差、すなわち、各 C_α 原子の位置の距離 (以下、単に偏差という。) l_{ii} を次の数7を用いて残基番号に対してプロットしたものである。

【0046】

【数7】 $l_{ii} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$

【0047】ここで、 (x_1, y_1, z_1) は目的蛋白質の3次元座標であり、 (x_2, y_2, z_2) は参照蛋白質の3次元座標である。図7から明らかなように、予測精度が高い領域では偏差 l_{ii} の値は小さく、予測精度の低い領域ではそれは大きくなっていることが認められる。

【0048】図8は、数6を用いて計算した相関関数の平均値 $\langle C(i) \rangle$ を残基番号に対してプロットしたものである。図7と比較することにより、目的蛋白質であるエラスターゼと参照蛋白質であるキモトリプシノーゲンとの間で相同性が高い、すなわち相関値が高い残基番号の領域では立体構造の予測精度が高く、すなわち偏差 l_{ii} の値が小さいが、一方、相同性が低い、すなわち相関値が低い残基番号の領域では立体構造の予測精度が低

く、すなわち偏差 l_{ii} の値が大きいことが認められる。このように、上記数6に基づいた相同性の定義に従えば、アミノ酸配列情報のみから予測された立体構造の局所的な予測精度の推定が可能となる。

【0049】以上の実施例において、立体構造の局所的な予測精度の推定を行うために、上記6つの物理化学的パラメータを用いているが、これらは本発明者が、立体構造が似ている部分に対応する残基番号の領域の相関値が高くなるようなパラメータを探した結果、いわば経験的に見つけたものである。本発明において用いる物理化学的パラメータとしては、好ましくは、上記6つの物理化学的パラメータのうちの少なくとも1つを用いるようにしてもよい。

【0050】本発明者は、上記の目的蛋白質であるエラスターゼと参照蛋白質であるキモトリプシノーゲンとの間の偏差 l_{ii} と相関関数の平均値 $\langle C(i) \rangle$ との関係のほかに、以下の目的蛋白質と参照蛋白質との間で、立体構造の局所的な予測精度の推定が可能であることを確認している。

(a) 目的蛋白質ヘモグロビン・アルファ鎖 (hemoglobin α -chain) と参照蛋白質ヘモグロビン・ベータ鎖 (hemoglobin β -chain)。

(b) 目的蛋白質アルファ・リステック・プロテアーゼ (α -lytic protease) と参照蛋白質プロテアーゼA (protease A)。

(c) 目的蛋白質エンドチアペプシン (endothiapepsin) と参照蛋白質ペニシロペプシン (penicillopepsin)。

【0051】次に、図1を参照して、本実施例の目的蛋白質の立体構造の予測演算及び予測精度演算装置の構成について説明する。図1に示すように、図2及び図3の処理を実行する演算制御装置であるマイクロプロセッシングユニット (以下、MPUという。) 10には、バスを介して以下の装置が接続される。

(a) ROM11: 図2及び図3のプログラムデータ及び当該プログラムを実行するために必要なデータを格納する。

(b) RAM12: 図2及び図3の処理を実行するためのワークエリアとなる記憶領域のためのメモリである。

(c) キーボード13: 図2及び図3の処理を実行するための指示コマンド又は目的蛋白質の名称と、アミノ酸配列と、並びに、予測座標計算処理における乱数のタネの値などを入力するための入力装置である。

(d) CRTディスプレイ14: 図2及び図3の処理の処理中の内容及び計算結果を表示する表示装置である。

(e) プリンタ15: 図2及び図3の処理の計算結果を印字する印字装置である。

(f) 計算結果ファイル20: 例えばハードディスクから構成され、図2及び図3の処理の計算結果のデータを記憶するための記憶装置である。

(g) 蛋白質データベース (PDB) ファイル 21 : 例
えばハードディスクから構成され、公知のX線結晶解析
及びNMRを用いた解析によって、種々の蛋白質を構成
するアミノ酸残基の各々に対してその空間座標 (x,
y, z) を求めた結果を含むデータベースファイルであ
って、本実施例において参照蛋白質となる得る各蛋白質
に関する蛋白質の名称、アミノ酸配列 (3文字コード表
示及び1文字コード表示)、残基数、各残基に対する空
間座標を記憶する記憶装置である。

【0052】図2は、図1の装置で実行される蛋白質の
立体構造予測及び精度計算処理を示すフローチャートで
ある。以下、目的蛋白質がエラストーゼであって参照蛋
白質がキモトリプシノーゲンであるときの一例を参照し
て、本実施例の蛋白質の立体構造予測及び精度計算処理
について説明する。なお、印字出力結果である表8乃至
表24については、明細書の編集の都合上、実施例の最
後に記載しかつ各表においてデータ長が1ページの長さ
を超えるものについては、複数の表に分割している。

【0053】図2において、まず、ステップS1におい
て、キーボード13を用いて、座標を予測したい目的蛋
白質の名称とアミノ酸配列とを入力した後、ステップS
2においてホモロジー検索処理が実行される。ホモロジ
ー検索処理においては、PDBに基づいて、目的蛋白質
のアミノ酸配列と類似する配列を有しかつ立体構造が既
知の参照蛋白質を検索し、対応する残基ペアが全体に占
める割合を示す相同性の度合いが例えば30%以上であ
る参照蛋白質候補のうち、本実施例では相同性の度合い
が最大である蛋白質を参照蛋白質に選択する。次いで、
ステップS3において、表8乃至表11に示すように、
以下の事項が印字される。

(a) 表8: 目的蛋白質の名称及び残基数

(b) 表9乃至表11: 上記選択された各参照蛋白質候
補について

(b-1) 参照蛋白質の名称

(b-2) 目的蛋白質と参照蛋白質との間の対応する残
基番号

(b-3) 相同性の度合い (%)

(b-4) 目的蛋白質と参照蛋白質との間のアミノ酸配
列の対応関係: ここで、アミノ酸は1文字コードで表さ
れ、上段が目的蛋白質であり下段が参照蛋白質であり、
残基番号の10毎に*が付されている。また、目的蛋白
質と参照蛋白質のアミノ酸が対応一致しているところに
“:” が付されており、アミノ酸が欠失しているところ
に“-” が付されている。

【0054】この例では、3つの参照蛋白質候補とし
て、[1] キモトリプシノーゲンAと、[2] トリプシ
ノーゲンと、[3] キモトリプシノーゲン*Aとが検索
され、本実施例では、これらの参照蛋白質候補のうち相
同性の度合いが最大であるキモトリプシノーゲン*A
(以下、単にキモトリプシノーゲンという。) が参照蛋

白質に選択される。

【0055】次いで、ステップS4において、選択され
た参照蛋白質 (キモトリプシノーゲン) と目的蛋白質
(エラストーゼ) との間のアミノ酸配列を比較して、本
実施例では、5残基以上の挿入又は欠失 (-) の無い対
応する相同な残基ペアを切り出して、ステップS5にお
いて、表12に示すように、対応する残基ペアの数と、
目的蛋白質と参照蛋白質の残基番号の対応表とが印字さ
れる。切り出される対応する残基ペアは、表11から明
らかなように以下の通りである。

(a) 目的蛋白質の残基番号1-23が参照蛋白質の残
基番号16-38に相同し、以下、残基ペアRP1とす
る。

(b) 目的蛋白質の残基番号28-53が参照蛋白質の
残基番号40-65に相同し、以下、残基ペアRP2と
する。

【0056】以下、同様に、残基ペアRP3乃至RP8
まで存在する。これら対応する残基ペアRP1乃至RP
8を図4に示す。本実施例では、図4に示すように、座
標が既知である参照蛋白質の残基番号 i' 番目の C_{α} 原
子の座標と残基番号 j' 番目の C_{α} 原子の座標との間の
相対距離を、それぞれ対応する目的蛋白質の残基番号 i
番目の C_{α} 原子の座標と残基番号 j 番目の C_{α} 原子の座
標との間の相対距離 $<d_{ij}>$ に設定して、目的蛋白質の
立体構造を予測する。さらに、ステップS6において、
図3の予測座標計算処理を実行して、目的蛋白質の C_{α}
原子の座標を計算して、計算結果をCRTディスプレイ
14に表示するとともに、プリンタ15を用いて印字す
る。

【0057】次いで、ステップS7において、上記20
種類のアミノ酸に固有の物理化学的パラメータ p の値に
基づいて、ある1つの物理化学的パラメータ p につい
て数5を用いて、目的蛋白質と参照蛋白質との間の相関関
数 $C_p(i)$ の値を計算し、次いで、さらに他の5種類
の物理化学的パラメータ p について同様に数5を用いて
目的蛋白質と参照蛋白質との間の相関関数 $C_p(i)$ の
値を計算し、すなわち6種類の物理化学的パラメータ p
について目的蛋白質と参照蛋白質との間の相関関数 C_p
(i) の値を計算した後、数6を用いて相関関数の平均
値 $<C(i)>$ を目的蛋白質の残基番号 i に対して計算
する。

【0058】さらに、ステップS8において、ステップ
S7において計算された相関関数の平均値 $<C(i)>$
を、図8のフォーマットで、すなわち目的蛋白質の残基
番号 i に対する相関関数の平均値 $<C(i)>$ を折れ線
グラフをCRTディスプレイ14に表示するとともに、
プリンタ15を用いて印字する。当該図8から、アミノ
酸配列情報のみに基づいて、予測された目的蛋白質の立
体構造の座標について残基番号に沿って1次元的に、局
所的な予測精度の推定が可能となった。この方法で1次

構造を比較し、立体構造の似ている可能性の高い部分を抽出することができる。当該方法によれば、相関関数の値がどれくらいであれば、また鎖がどれだけ長ければ立体構造の一致が期待されるかを検討することができる。

【0059】図3は、図2のサブルーチンである予測座標計算処理（ステップS6）を示すフローチャートである。図3に示すように、まず、ステップS11において、ステップS4において切り出された残基ベアRP1乃至RP8に基づいて、参照蛋白質における対応する相同な領域間の相対距離 d_{ij} を計算して、これを目的蛋白質の対応する相対距離 d_{ij} として設定し、この処理を $i < j$ の条件のもとで i と j に関して繰り返し実行する。次いで、ステップS12において、乱数のタネIRを入力する。ステップS13においては、入力した乱数のタネIRに基づいて乱数値（ $0 \leq \text{乱数値} < 1$ ）を発生し、さらに、上記数4を用いて、目的蛋白質の初期構造の座標を実際の座標により近似させるために、上記乱数値に蛋白質の旋回半径 R_g をかけた後、直径の座標とするためにそれを2倍して目的蛋白質の初期構造の座標 q_{ij} を計算しそれに基づいて上記初期構造の相対距離 d_{ij} を計算し、この計算を $i < j$ の条件のもとで i と j に関して繰り返し実行する。そして、ステップS14において当該初期構造における後述する数1の関数値Fを計算する。

【0060】次いで、ステップS15において、例えば1階の偏微分のみを計算するが2階の偏微分を計算しないで関数の最小化を行う公知の準ニュートン法を用いて、数1の目的蛋白質の関数Fの関数値が最小になるように関数Fの最適化を行い、関数値Fを計算する。さらに、ステップS16において、最適化された関数Fに基づいて、関数Fの変数である目的蛋白質の予測座標変数 $x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_M, y_M, z_M$ を計算する。なお、ステップS14、S15において目的蛋白質の初期構造における関数値Fと最適化後の目的蛋白質の関数値Fとを計算するのは、最適化の度合いを知るためである。最適化後の目的蛋白質の関数値Fが目的蛋白質の初期構造における関数値Fに比較して十分に小さくなったとき、最適化の度合いが大きいと判定することができる。

【0061】さらに、ステップS17において、表13と表14に示す標準出力フォーマットで出力結果を印字した後、ステップS18において、表15乃至表24に示す目的蛋白質の予測座標の計算結果を印字する。

【0062】上記標準出力フォーマットにおいては、表13と表14に示すように、以下の情報が印字される。

(a) 目的蛋白質の名称及びアミノ酸配列（1文字コード表示）

(b) 参照蛋白質の名称及びアミノ酸配列（3文字コード表示）

(c) 目的蛋白質と参照蛋白質との間で対応する相同な

残基番号

(d) 乱数のタネの値

(e) ステップS15における最適化における最適化の計算の繰り返し数

(f) ステップS14で計算された目的蛋白質の初期構造の関数値F

(g) ステップS15で計算された最適化後の目的蛋白質の予測構造の関数値F

【0063】さらに、上記予測座標の計算結果においては、表15乃至表24に示すように、以下の情報が印字される。

(a) 目的蛋白質の名称

(b) 目的蛋白質の残基毎の以下のデータ

(b-1) アミノ酸配列（3文字コード表示）

(b-2) 残基番号

(b-3) 予測された x, y, z 座標値

【0064】以上説明したように、本実施例によれば、PDBに基づいて相同性の度合いが最大又は比較的高い蛋白質を参照蛋白質を選択し、選択された参照蛋白質と目的蛋白質との間のアミノ酸配列を比較して、相同な複数の残基ベアを切り出し、切り出された複数の残基ベアに基づいて、参照蛋白質における互いに残基番号が異なる各残基の原子間の相対距離を計算して、それを目的蛋白質の対応する相対距離 d_{ij} として設定し、所定の乱数のタネIRに基づいて乱数を発生させて乱数値に比例する値を目的蛋白質の初期構造の座標値として設定して初期構造の相対距離 d_{ij} を計算した後、目的蛋白質の関数Fの関数値が最小になるように関数Fの最適化を行い、関数Fの変数である目的蛋白質の予測座標を計算している。従って、従来の方法に比較して計算をより簡単にすることができるとともに、輪湖・シェラガの方法などの従来の方法に比較して目的蛋白質の座標をより高い精度で、すなわちより実際に近い座標を計算することができる。

【0065】さらに、目的蛋白質の残基番号 i に対する相関関数の平均値 $\langle C(i) \rangle$ を上記数5と数6とを用いて計算しかつ表示することにより、アミノ酸配列情報のみに基づいて、予測された目的蛋白質の立体構造の座標について残基番号に沿って1次元的に、局所的な予測精度の推定が可能となる。この方法で1次構造を比較し、立体構造の似ている可能性の高い部分を抽出することができる。当該方法によれば、相関関数の値がどれくらいであれば、また鎖がどれだけ長ければ立体構造の一致が期待されるかを検討することができる。

【0066】

【表8】

Target protein:

elastase porcine ec 3.4.21.36

No. of residues: 240

Fragmentation from 1 to 240

【0067】

* * 【表9】

[1]

(1 - 240) 240 elastase porcine ec 3.4.21.36

(16 - 226) 226 CHYMOTRYPSINOGEN A

code name: 1CHG cor. coef.= 0.415 (37.9%)

```

+ * + * + * + * + * + * + *
VVGTEAQRNSWPSQISLQYRSGSSWAHTCGGTLIRQNWMTAAHCVDRELTFRVVVGHEHNLNQNGTEQ
: : : : : : : : : : : : : : : :
IVNGEEAVPGSWPWQVSLQDKTG-F-HFCGGLINENWVVTAAHC-GVT-TSDVVVAGEIKQLK-IAK
+ * + * + * + * + * + * + *
YVGQKIVVHPYWNDDVAAGYDIALRLAQSVTLNSYVQLGVLPRAGTILANNSPCYITGWLTRTNGQ
: : : : : : : : : : : : : : : :
-V-F-KNS-K-Y-NSLTINN-DITLLKLSTAASFSTVSAVCLPSASDDFAAGTTCVITGWG-R-
+ * + * + * + * + * + * + *
LAQTLQQAAYLPTVDYAI CSSSSSYWGSTVKNSMVCAGNGVRSQCQDSCGGLHCLVNGQYAVHGVTSFVS
: : : : : : : : : : : : : : : :
L-Q-QAS-LPLLSNTNCKK-YWGTRIKDAMICAGASGV-SSCMGDSGGPLVCKKNGAWTLVGIVSWGS
+ * + * + * + * + * + * + *
RLGCNVTRKPTVFTRVSAVISWINNVIASN
: : : : : : : : : : : : : : : :
S-TCS-TSTPGVYARVTALVNWVQQLAAN

```

【0068】

* * 【表10】

[2]

(17 - 240) 240 elastase porcine ec 3.4.21.36

(10 - 222) 222 TRYPSINOGEN

code name: 1TGN cor. coef.= 0.452 (35.7%)

```

+ * + * + * + * + * + * + *
SLQYRSGSSWA-HTCGGTLIRQNWMTAAHCVDRELTFRVVVGHEHNLNQNGTEQYVGQKIVVHPYWN
: : : : : : : : : : : : : : : :
TVPYQVSLNSGYHFCGGLINSQWVVSAAHCY-KS-GIQVRLGEDNINVVEGNEQFISAKSIVHPSYNS
+ * + * + * + * + * + * + *
DDVAAGYDIALRLAQSVTLNSYVQLGVLPRAGTILANNSPCYITGWLTRTNGQ-LAQTLQQAAYLPTVD
: : : : : : : : : : : : : : : :
NTLNN-DIMLIKLSAASLSNRVASISLP-TSCA-SAGTQCLISGWNTKSSGTSYPDLKCLKAPILS
+ * + * + * + * + * + * + *
YAI CSSSSSYWGSTVKNSMVCAG-GNGVRSQCQDSCGGLHCLVNGQYAVHGVTSFVSRLGCNVTRKPTVF
: : : : : : : : : : : : : : : :
DSSCKSA-Y-PGQITSNMFCAGYLEGGKDSQCQDGGPVVC-SGK-LQGI VSWGS-GCAQKNKPGVY
+ * + * + * + * + * + * + *
TRVSAYISWINNVIASN
: : : : : : : : : : : : : : : :
TKVCNVVSWIKQTIASN

```

【0069】

★ ★ 【表11】

[3]

(1 - 240) 240 elastase porcine ec 3.4.21.36

(16 - 245) 245 CHYMOTRYPSINOGEN *A

code name: 2CGA cor. coef.= 0.502 (39.0%)

```

+ * + * + * + * + * + * + *
VVGTEAQRNSWPSQISLQYRSGSSWAHTCGGTLIRQNWMTAAHCVDRELTFRVVVGHEHNLNQNGTEQ
: : : : : : : : : : : : : : : :

```

20

*

※ ※【表14】

CYS GLY VAL PRO ALA ILE GLN PRO VAL LEU SER GLY LEU SER ARG ILE VAL ASN
GLY GLU GLU ALA VAL PRO GLY SER TRP PRO TRP GLN VAL SER LEU GLN ASP LYS
THR GLY PHE HIS PHE CYS GLY GLY SER LEU ILE ASN GLU ASN TRP VAL VAL THR
ALA ALA HIS CYS GLY VAL THR THR SER ASP VAL VAL VAL ALA GLY GLU PHE ASP
GLN GLY SER SER SER GLU LYS ILE GLN LYS LEU LYS ILE ALA LYS VAL PHE LYS
ASN SER LYS TYR ASN SER LEU THR ILE ASN ASN ASP ILE THR LEU LEU LYS LEU
SER THR ALA ALA SER PHE SER GLN THR VAL SER ALA VAL CYS LEU PRO SER ALA
SER ASP ASP PHE ALA ALA GLY THR THR CYS VAL THR THR GLY TRP GLY LEU THR
ARG TYR THR ASN ALA ASN THR PRO ASP ARG LEU GLN GLN ALA SER LEU PRO LEU
LEU SER ASN THR ASN CYS LYS LYS TYR TRP GLY THR LYS ILE LYS ASP ALA MET
ILE CYS ALA GLY ALA SER GLY VAL SER SER CYS MET GLY ASP SER GLY GLY PRO
LEU VAL CYS LYS LYS ASN GLY ALA TRP THR LEU VAL GLY ILE VAL SER TRP GLY

23

24

SER SER THR CYS SER THR SER THR PRO GLY VAL TYR ALA ARG VAL THR ALA LEU
VAL ASN TRP VAL GLN GLN THR LEU ALA ALA ASN

Residue pairs:

1- 23: 16- 38 | 28- 53: 40- 65 | 55- 81: 66- 92 | 84-136: 93-145 |
137-160:147-170 | 163-180:171-188 | 182-211:189-218 | 217-240:222-245 |

Random number generation: 2

Minop has exhausted 300calls of calcfg

** Initial conformation: F= 0.521134E+07

** Final conformation: F= 0.323426E+04

[0073]

10 [表15]

HEADER	elastase	porcine	ec	3.4.21.36
COMPND				
ATOM	1 CA VAL	1	18.851	-2.532 20.987
ATOM	2 CA VAL	2	15.303	-2.327 19.641
ATOM	3 CA GLY	3	14.268	1.317 20.378
ATOM	4 CA GLY	4	16.678	2.285 23.284
ATOM	5 CA THR	5	15.808	2.957 26.986
ATOM	6 CA GLU	6	14.409	5.605 29.237
ATOM	7 CA ALA	7	16.874	8.289 30.295
ATOM	8 CA GLN	8	17.404	9.265 33.926
ATOM	9 CA ARG	9	15.443	12.444 34.524
ATOM	10 CA ASN	10	17.432	15.682 34.242
ATOM	11 CA SER	11	20.388	13.904 32.727
ATOM	12 CA TRP	12	20.195	15.463 29.240
ATOM	13 CA PRO	13	19.813	19.086 30.382
ATOM	14 CA SER	14	20.126	20.787 26.964
ATOM	15 CA GLN	15	17.127	18.902 25.531
ATOM	16 CA ILE	16	14.106	21.076 24.752
ATOM	17 CA SER	17	10.689	20.574 23.259
ATOM	18 CA LEU	18	9.093	22.827 20.699
ATOM	19 CA GLN	19	5.414	23.298 20.964
ATOM	20 CA TYR	20	2.907	25.510 19.275
ATOM	21 CA ARG	21	0.476	28.157 20.613
ATOM	22 CA SER	22	-2.005	25.426 21.582
ATOM	23 CA GLY	23	0.535	22.932 22.760
ATOM	24 CA SER	24	0.467	21.756 19.096
ATOM	25 CA SER	25	2.042	22.804 15.793

[0074]

* * [表16]

ATOM	26 CA TRP	26	3.969	19.750 14.634
ATOM	27 CA ALA	27	2.807	17.611 17.585
ATOM	28 CA HIS	28	4.789	18.927 20.563
ATOM	29 CA THR	29	6.194	19.127 17.020
ATOM	30 CA CYS	30	9.957	19.301 17.210
ATOM	31 CA GLY	31	12.964	19.010 19.458
ATOM	32 CA GLY	32	15.944	21.369 20.022
ATOM	33 CA THR	33	19.182	21.917 22.057
ATOM	34 CA LEU	34	20.207	24.855 24.311
ATOM	35 CA ILE	35	23.569	26.252 23.416
ATOM	36 CA ARG	36	23.218	28.969 26.009
ATOM	37 CA GLN	37	20.595	30.780 28.048

25

26

【0075】

ATOM	38	CA	ASN	38	19.501	32.806	24.995
ATOM	39	CA	TRP	39	19.762	30.486	22.041
ATOM	40	CA	VAL	40	18.339	27.129	20.770
ATOM	41	CA	MET	41	19.501	25.071	17.786
ATOM	42	CA	THR	42	16.988	23.035	15.819
ATOM	43	CA	ALA	43	16.242	21.851	12.273
ATOM	44	CA	ALA	44	15.217	24.288	9.484
ATOM	45	CA	HIS	45	12.578	21.796	8.167
ATOM	46	CA	CYS	46	10.856	22.238	11.555
ATOM	47	CA	VAL	47	9.278	25.470	10.248
ATOM	48	CA	ASP	48	9.424	27.290	13.625
ATOM	49	CA	ARG	49	7.681	30.673	14.011
ATOM	50	CA	GLU	50	7.681	33.203	16.785

* * 【表17】

ATOM	51	CA	LEU	51	4.274	31.867	17.943
ATOM	52	CA	THR	52	5.811	28.550	18.988
ATOM	53	CA	PHE	53	7.071	28.006	22.591
ATOM	54	CA	ARG	54	8.397	24.718	21.106
ATOM	55	CA	VAL	55	10.286	26.413	23.928
ATOM	56	CA	VAL	56	9.932	24.014	26.848
ATOM	57	CA	VAL	57	12.973	23.250	28.996
ATOM	58	CA	GLY	58	13.598	20.851	31.890
ATOM	59	CA	GLU	59	11.025	18.298	30.786
ATOM	60	CA	HIS	60	11.118	14.540	31.293
ATOM	61	CA	ASN	61	7.653	13.062	31.184
ATOM	62	CA	LEU	62	5.258	14.966	29.043
ATOM	63	CA	ASN	63	2.358	13.384	30.844
ATOM	64	CA	GLN	64	3.344	15.002	34.151
ATOM	65	CA	ASN	65	2.043	18.438	35.406
ATOM	66	CA	ASN	66	4.203	18.063	38.430
ATOM	67	CA	GLY	67	7.441	18.996	36.718
ATOM	68	CA	THR	68	9.139	22.298	37.292
ATOM	69	CA	GLU	69	9.719	23.150	33.610
ATOM	70	CA	GLN	70	10.445	26.522	31.865
ATOM	71	CA	TYR	71	8.236	27.651	29.000
ATOM	72	CA	VAL	72	10.232	30.118	26.973
ATOM	73	CA	GLY	73	9.090	32.602	24.329
ATOM	74	CA	VAL	74	10.871	33.193	21.048
ATOM	75	CA	GLN	75	12.145	36.637	19.897

【0076】

※40※ 【表18】

ATOM	76	CA	LYS	76	13.419	35.909	16.359
ATOM	77	CA	ILE	77	13.963	32.885	14.160
ATOM	78	CA	VAL	78	17.058	32.573	11.944
ATOM	79	CA	VAL	79	16.921	29.883	9.269
ATOM	80	CA	HIS	80	20.294	29.135	7.395
ATOM	81	CA	PRO	81	19.837	30.674	3.980
ATOM	82	CA	TYR	82	16.953	28.351	4.957
ATOM	83	CA	TRP	83	20.573	27.735	5.984
ATOM	84	CA	ASN	84	21.686	27.864	2.292
ATOM	85	CA	THR	85	18.733	25.451	3.256

ATOM	86	CA	ASP	86	17.599	23.567	0.387
ATOM	87	CA	ASP	87	14.123	22.580	0.848
ATOM	88	CA	VAL	88	14.430	20.145	-2.002
ATOM	89	CA	ALA	89	17.363	18.113	-0.848
ATOM	90	CA	ALA	90	16.890	18.918	2.901
ATOM	91	CA	GLY	91	20.567	19.922	2.948
ATOM	92	CA	TYR	92	21.984	22.543	5.411
ATOM	93	CA	ASP	93	19.056	21.731	7.820
ATOM	94	CA	ILE	94	19.470	24.142	10.697
ATOM	95	CA	ALA	95	17.722	27.016	12.452
ATOM	96	CA	LEU	96	18.585	29.300	15.403
ATOM	97	CA	LEU	97	16.026	30.577	17.739
ATOM	98	CA	ARG	98	16.756	33.629	19.888
ATOM	99	CA	LEU	99	14.620	33.510	23.091
ATOM	100	CA	ALA	100	12.579	36.612	24.198

【0077】

* * 【表19】

ATOM	101	CA	GLN	101	12.921	35.234	27.696
ATOM	102	CA	SER	102	16.256	33.781	28.839
ATOM	103	CA	VAL	103	16.598	30.229	30.219
ATOM	104	CA	THR	104	17.429	29.900	33.841
ATOM	105	CA	LEU	105	20.290	27.410	34.114
ATOM	106	CA	ASN	106	19.975	24.821	36.900
ATOM	107	CA	SER	107	20.655	21.250	37.796
ATOM	108	CA	TYR	108	18.309	20.142	35.019
ATOM	109	CA	VAL	109	18.669	22.991	32.486
ATOM	110	CA	GLN	110	22.024	23.750	30.864
ATOM	111	CA	LEU	111	23.957	23.834	27.556
ATOM	112	CA	GLY	112	25.474	21.338	25.072
ATOM	113	CA	VAL	113	28.934	22.164	23.644
ATOM	114	CA	LEU	114	29.414	23.002	19.950
ATOM	115	CA	PRO	115	32.334	21.340	17.893
ATOM	116	CA	ARG	116	35.074	23.456	16.017
ATOM	117	CA	ALA	117	34.676	23.489	12.202
ATOM	118	CA	GLY	118	37.846	21.310	12.129
ATOM	119	CA	THR	119	36.637	18.684	14.544
ATOM	120	CA	ILE	120	36.853	15.237	13.204
ATOM	121	CA	LEU	121	34.484	12.657	14.608
ATOM	122	CA	ALA	122	35.483	9.192	13.197
ATOM	123	CA	ASN	123	33.124	6.458	11.904
ATOM	124	CA	ASN	124	32.277	3.845	14.454
ATOM	125	CA	SER	125	32.462	6.308	17.289

【0078】

※ ※ 【表20】

ATOM	126	CA	PRO	126	29.443	5.603	19.607
ATOM	127	CA	CYS	127	27.161	8.681	20.144
ATOM	128	CA	TYR	128	23.801	9.364	21.797
ATOM	129	CA	ILE	129	20.457	10.524	20.417
ATOM	130	CA	THR	130	17.632	11.463	22.732
ATOM	131	CA	GLY	131	13.996	12.472	22.429
ATOM	132	CA	TRP	132	10.244	11.899	22.536
ATOM	133	CA	GLY	133	10.010	10.720	18.922

ATOM	134	CA	LEU	134	8.160	7.549	19.898
ATOM	135	CA	THR	135	5.145	9.670	20.637
ATOM	136	CA	ARG	136	4.723	10.340	16.882
ATOM	137	CA	THR	137	-0.358	8.075	17.391
ATOM	138	CA	ASN	138	-2.041	4.683	17.584
ATOM	139	CA	GLY	139	0.642	3.241	19.679
ATOM	140	CA	GLN	140	0.073	6.394	21.641
ATOM	141	CA	LEU	141	3.334	5.402	23.381
ATOM	142	CA	ALA	142	5.564	6.033	26.461
ATOM	143	CA	GLN	143	5.945	9.802	26.947
ATOM	144	CA	THR	144	9.266	9.810	28.849
ATOM	145	CA	LEU	145	12.605	10.935	27.345
ATOM	146	CA	GLN	146	14.358	7.949	25.590
ATOM	147	CA	GLN	147	18.018	7.543	24.736
ATOM	148	CA	ALA	148	19.880	5.281	22.299
ATOM	149	CA	TYR	149	23.601	4.723	21.595
ATOM	150	CA	LEU	150	24.381	4.597	17.895

【0079】

* * 【表21】

ATOM	151	CA	PRO	151	27.725	4.413	15.936
ATOM	152	CA	THR	152	28.655	7.144	13.428
ATOM	153	CA	VAL	153	28.940	6.048	9.799
ATOM	154	CA	ASP	154	31.047	7.478	6.999
ATOM	155	CA	TYR	155	28.950	9.044	4.203
ATOM	156	CA	ALA	156	30.379	6.460	1.703
ATOM	157	CA	ILE	157	28.888	3.802	3.705
ATOM	158	CA	CYS	158	25.671	5.686	4.243
ATOM	159	CA	SER	159	25.080	5.897	0.451
ATOM	160	CA	SER	160	24.542	2.092	0.480
ATOM	161	CA	SER	161	23.473	1.851	4.162
ATOM	162	CA	SER	162	20.506	3.770	5.660
ATOM	163	CA	TYR	163	21.189	2.940	2.015
ATOM	164	CA	TRP	164	20.008	6.191	0.847
ATOM	165	CA	GLY	165	22.043	6.686	-2.295
ATOM	166	CA	SER	166	22.599	10.145	-3.735
ATOM	167	CA	THR	167	20.332	11.720	-1.147
ATOM	168	CA	VAL	168	23.467	11.834	1.098
ATOM	169	CA	LYS	169	25.090	15.334	0.833
ATOM	170	CA	ASN	170	28.351	16.462	2.426
ATOM	171	CA	SER	171	26.602	18.580	5.035
ATOM	172	CA	MET	172	24.920	15.473	6.321
ATOM	173	CA	VAL	173	26.021	12.759	8.769
ATOM	174	CA	CYS	174	24.491	9.298	9.375
ATOM	175	CA	ALA	175	24.378	7.183	12.460

【0080】

※ ※ 【表22】

ATOM	176	CA	GLY	176	22.974	3.824	13.593
ATOM	177	CA	GLY	177	21.787	0.767	11.650
ATOM	178	CA	ASN	178	23.430	-1.515	14.182
ATOM	179	CA	GLY	179	20.439	-3.047	15.884
ATOM	180	CA	VAL	180	18.859	-0.111	17.378
ATOM	181	CA	ARG	181	19.056	3.286	18.939

ATOM	182	CA	SER	182	16.589	2.327	16.140
ATOM	183	CA	GLY	183	16.885	6.146	15.520
ATOM	184	CA	CYS	184	13.181	7.048	15.645
ATOM	185	CA	GLN	185	13.352	8.263	19.312
ATOM	186	CA	GLY	186	15.002	11.524	18.037
ATOM	187	CA	ASP	187	12.469	14.253	17.449
ATOM	188	CA	SER	188	13.025	16.375	14.353
ATOM	189	CA	GLY	189	15.565	18.979	15.251
ATOM	190	CA	GLY	190	16.815	17.051	18.336
ATOM	191	CA	PRO	191	20.443	16.249	19.282
ATOM	192	CA	LEU	192	23.005	13.724	18.326
ATOM	193	CA	HIS	193	25.791	14.108	20.861
ATOM	194	CA	CYS	194	29.203	12.596	21.186
ATOM	195	CA	LEU	195	31.753	13.081	23.894
ATOM	196	CA	VAL	196	34.846	15.124	23.079
ATOM	197	CA	ASN	197	37.452	15.597	25.869
ATOM	198	CA	GLY	198	34.806	14.392	28.396
ATOM	199	CA	GLN	199	32.250	16.930	27.291
ATOM	200	CA	TYR	200	29.016	16.171	25.402

【0081】

* 20 * 【表 2 3】

ATOM	201	CA	ALA	201	29.237	17.882	22.017
ATOM	202	CA	VAL	202	26.477	18.407	19.386
ATOM	203	CA	HIS	203	27.599	16.495	16.297
ATOM	204	CA	GLY	204	24.149	15.989	14.545
ATOM	205	CA	VAL	205	20.649	17.525	14.312
ATOM	206	CA	THR	206	17.833	14.988	13.502
ATOM	207	CA	SER	207	17.004	15.670	9.868
ATOM	208	CA	PHE	208	15.361	12.670	8.212
ATOM	209	CA	VAL	209	15.198	8.921	7.541
ATOM	210	CA	SER	210	12.840	6.012	7.980
ATOM	211	CA	ARG	211	9.065	6.721	8.446
ATOM	212	CA	LEU	212	9.207	10.363	9.487
ATOM	213	CA	GLY	213	10.513	8.176	12.286
ATOM	214	CA	CYS	214	14.299	8.540	12.751
ATOM	215	CA	ASN	215	14.022	5.057	14.226
ATOM	216	CA	VAL	216	16.089	2.589	12.218
ATOM	217	CA	THR	217	16.970	-0.988	13.315
ATOM	218	CA	ARG	218	18.852	-1.719	10.187
ATOM	219	CA	LYS	219	18.625	1.675	8.439
ATOM	220	CA	PRO	220	20.773	4.709	9.586
ATOM	221	CA	THR	221	19.186	7.973	10.532
ATOM	222	CA	VAL	222	20.240	11.136	8.710
ATOM	223	CA	PHE	223	21.377	14.250	10.543
ATOM	224	CA	THR	224	22.588	17.709	9.840
ATOM	225	CA	ARG	225	26.434	17.861	10.202

【0082】

※ ※ 【表 2 4】

ATOM	226	CA	VAL	226	26.987	20.558	12.851
ATOM	227	CA	SER	227	30.732	20.894	12.035
ATOM	228	CA	ALA	228	29.767	22.035	8.475
ATOM	229	CA	TYR	229	27.315	24.546	9.917

33

34

ATOM	230	CA	ILE	230	29.176	25.852	12.929
ATOM	231	CA	SER	231	30.676	28.924	11.237
ATOM	232	CA	TRP	232	27.267	30.374	10.371
ATOM	233	CA	ILE	233	26.138	29.758	14.035
ATOM	234	CA	ASN	234	29.038	31.790	15.523
ATOM	235	CA	ASN	235	28.410	34.403	12.987
ATOM	236	CA	VAL	236	24.670	34.745	13.693
ATOM	237	CA	ILE	237	25.176	34.793	17.433
ATOM	238	CA	ALA	238	28.014	37.384	17.237
ATOM	239	CA	SER	239	25.674	39.775	15.407
ATOM	240	CA	ASN	240	22.208	39.280	16.899

END

【0083】以上の実施例において、目的蛋白質の α 炭素原子の座標を予測演算しているが、本発明はこれに限らず、同様に、蛋白質を構成するすべての原子であるアミノ酸残基の原子の座標を予測演算するように構成してもよい。

【0084】

【発明の効果】以上詳述したように本発明によれば、アミノ酸残基の原子の座標が既知の蛋白質のうち、立体構造を予測すべき目的蛋白質との相溶性の度合いが高い少なくとも1つの蛋白質を参照蛋白質に選択し、上記選択した参照蛋白質のアミノ酸残基の原子の座標に基づいて上記目的蛋白質のアミノ酸残基の原子の座標が所定の方法で予測された目的蛋白質の立体構造の予測精度を演算する蛋白質の立体構造の予測精度演算方法及び装置であって、上記目的蛋白質に含まれる複数種類のアミノ酸の物理化学的パラメータに基づいて、上記参照蛋白質と上記目的蛋白質の併置配列において各対応するアミノ酸残基に対する相溶性を示す相関関数の値を、上記複数種類のアミノ酸の物理化学的パラメータから予め選択された少なくとも1つの選択物理化学的パラメータについてそれぞれ計算し、上記計算した相関関数の値を平均して計算された相関関数の平均値を上記予測された目的蛋白質の立体構造の予測精度として用いる。従って、本発明は、アミノ酸配列情報のみに基づいて、予測された目的蛋白質の立体構造の座標について残基番号に沿って1次元的に、局所的な予測精度の推定が可能となる。この方法で1次構造を比較し、立体構造の似ている可能性の高い部分を抽出することができる。当該方法によれば、相関関数の値がどれくらいであれば、また鎖がどれだけ長ければ立体構造の一致が期待されるかを検討することができる。

【図面の簡単な説明】

【図1】 本発明に係る一実施例である蛋白質の立体構

造の予測演算及び予測精度演算装置の構成を示すブロック図である。

【図2】 図1の装置で実行される蛋白質の立体構造予測及び精度計算処理を示すフローチャートである。

【図3】 図2のサブルーチンである予測座標計算処理を示すフローチャートである。

【図4】 目的蛋白質がエラストーゼであり参照蛋白質がキモトリプシノーゲンであるときの相溶性残基ペアと本発明に係る相対距離 $\langle d_{ij} \rangle$ を示す図である。

【図5】 本発明で用いる目的蛋白質Xの残基番号*i*と参照蛋白質Yの残基番号*j*との間の相関関数 $C_p(i)$ を示す数5を説明するための図である。

【図6】 図1の装置で予測された目的蛋白質エラストーゼの立体構造(実線)とそのX線結晶解析による計測結果の立体構造(破線)とを示すステレオ図である。

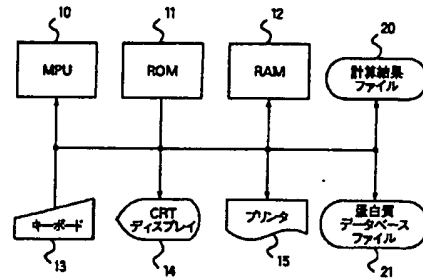
【図7】 目的蛋白質エラストーゼの残基番号に対する、図1の装置で予測された目的蛋白質エラストーゼの立体構造とそのX線結晶解析による計測結果の立体構造との間の偏差 l_{ij} を示すグラフである。

【図8】 目的蛋白質エラストーゼの残基番号に対する、図1の装置で計算された予測精度係数を示す相関関数 $C_p(i)$ の平均値 $\langle C(i) \rangle$ を示すグラフである。

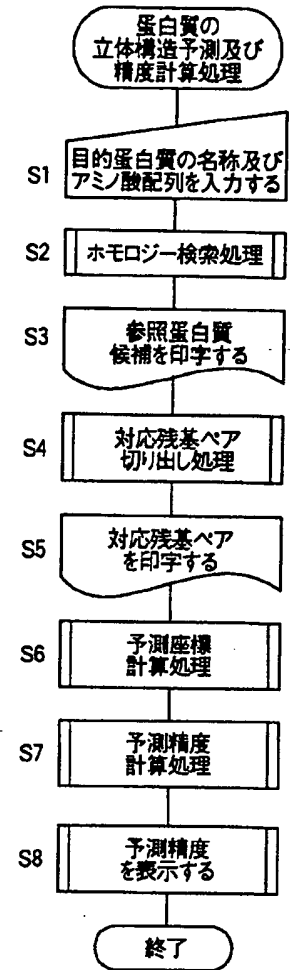
【符号の説明】

- 10…マイクロプロセッシングユニット(MPU)、
- 11…ROM、
- 12…RAM、
- 13…キーボード、
- 14…CRTディスプレイ、
- 15…プリンタ、
- 20…計算結果ファイル、
- 21…蛋白質データベース(PDB)ファイル、
- RP1乃至RP8…対応する残基ペア。

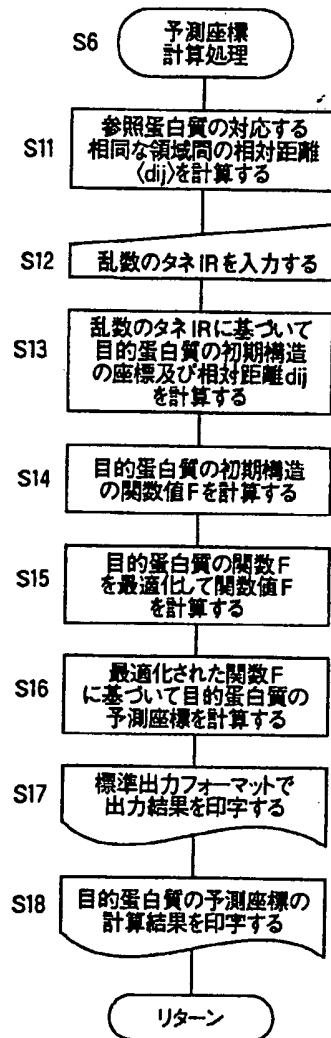
【図1】



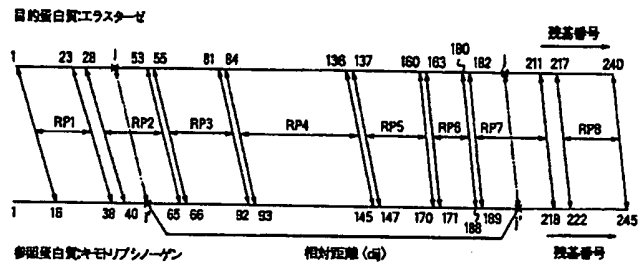
【図2】



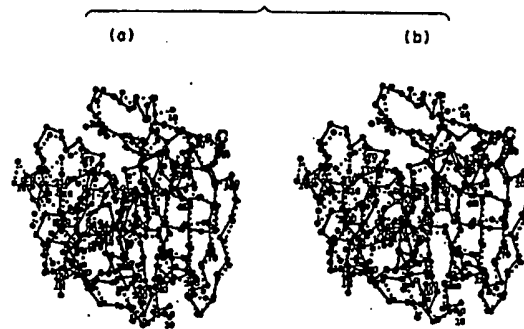
【図3】



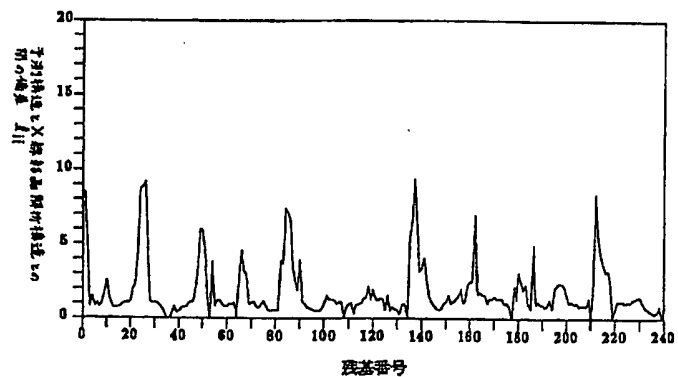
【図4】



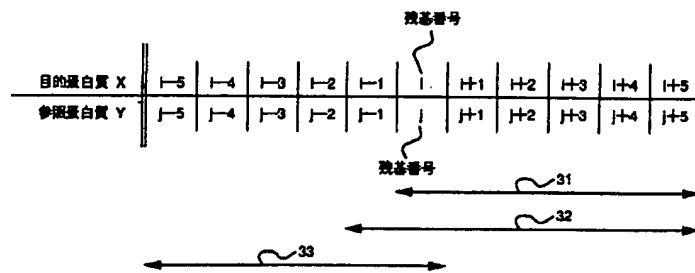
【図6】



【図7】



【図5】



【図8】

